ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 12, Nomor 3, Tahun 2023, Halaman 330 - 339
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

KAJIAN SIMULASI PERBANDINGAN METODE *RIDGE REGRESSION* DAN *ADJUSTED RIDGE REGRESSION* UNTUK PENANGANAN MULTIKOLINEARITAS

Choirun Nisa*, Siti Hariati Hastuti[2]
[1] Fakultas Ekonomi dan Sosial, Universitas Amikom Yogyakarta
[2]Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hamzanwadi
***e-mail**: choirun.nisa@amikom.ac.id

DOI: 10.14710/j.gauss.12.3.330-339

Article Info:
Received: 2023-01-29
Accepted: 2023-09-19
Available Online: 2024-02-13

Keywords:
multicollinearity; ridge regression; adjusted ridge regression; MSE.

Abstract: Regression analysis is widely used in research. However, often in using this analysis the assumption of non-multicollinearity is not fulfilled. Handling of these problems can be done using Ridge Regression (RR) and Adjusted Ridge Regression (AR) methods. This study aims to compare the performance of RR and AR in handling multicollinearity among explanatory variables in multiple regression analysis using data simulation. The simulated data contain various multicollinearity level (ρ = 0.6, 0.8, 0.9) with of each different sample size (n = 20, 50, 100). The performance of the two methods are compared using Mean Square Errors (MSE). The result shows that the AR method and the RR method produce a smaller MSE value when the sample size used is larger. The MSE value generated by the AR method tends to be smaller than the RR method which can be seen from each data repetition used. It shows that the AR method is relatively more effective than the RR method for dealing with multicollinearity problems.

1. PENDAHULUAN

Salah satu analisis statistika yang populer dan sering digunakan dalam penelitian serta pengambilan keputusan adalah analisis regresi. Analisis regresi dapat dimanfaatkan untuk memperkirakan sejauh mana variabel prediktor mempengaruhi variabel respon (Iqbal, 2021). Secara lebih rinci, Iriawan dan Astuti (2006) menyebutkan kegunaan analisis regresi dalam penelitian, antara lain: (1) mengukur kekuatan hubungan antara variabel respon dan variabel prediktor; (2) mengetahui pengaruh suatu atau beberapa variable prediktor terhadap variabel respon; serta (3) memprediksi pengaruh suatu atau beberapa variabel prediktor terhadap variabel respon.

Terdapat beberapa asumsi yang harus dipenuhi pada analisis regresi linier berganda, yaitu normalitas residual, homoskesdastisitas, tidak adanya autokorelasi, dan tidak adanya multikolinieritas antar variabel prediktor. Multikolinieritas merupakan masalah yang sering terjadi dalam analisis regresi. Kondisi ini mengakibatkan determinan dari matriks mendekati nol sehingga matriks tersebut hampir *singular* yang berakibat terhadap nilai penduga parameter tidak stabil. Menurut Montgomery dan Peck (1992), dampak multikolinieritas mengakibatkan koefisien regresi menjadi sangat lemah atau tidak dapat memberikan hasil analisis yang mewakili sifat dari variabel prediktor bersangkutan. Daoud (2017) lebih jauh menjelaskan bahwa kondisi ini mengakibatkan parameter model regresi kemudian tidak dapat diinterpretasikan.

Hoerl dan Kennard (1970) memperkenalkan metode untuk menangani kasus multikolinearitas pada regresi berganda yang disebut sebagai *Ridge Regression* (RR). Metode ini merupakan suatu teknik yang dikembangkan untuk menstabilkan nilai koefisien regresi dengan menentukan konstanta bias (k) namun memiliki keragaman lebih kecil. Beberapa peneliti yang telah menggunakan metode ini di antaranya Hildawati *et al*. (2016) dengan studi kasus upah minimum; Tazliqoh *et al*. (2015) dengan studi kasus pendapatan asli daerah; dan Sulistianingsih *et al*. (2023) dengan studi kasus indeks pembangunan manusia.

Banyak peneliti yang telah mengusulkan perkiraan berbeda untuk pemilihan konstanta tersebut, sehingga belum terdapat formula RR yang secara ekspilist dan tunggal. Terdapat estimasi alternatif lain yang diperkenalkan oleh Dorugade (2016) yaitu *Adjusted Ridge Regression* (AR) yakni sebuah modifikasi dari RR dengan mengubah konstanta bias dengan memanfaatkan nilai dari vektor. Metode AR dapat menangani kasus dengan multikolinieritas tinggi atau ekstrem pada regresi linier berganda. Penelitian ini bertujuan untuk menggunakan metode AR dan perbandingannya dengan metode RR dengan melihat nilai *Mean Square Error* (MSE) melalui data simulasi.

2. TINJAUAN PUSTAKA

Regresi *ridge* mulai diperkenalkan oleh Hoerl dan Kennard (1970). Metode regresi *ridge* bertujuan untuk mengatasi multikolinieritas yang terjadi pada regresi berganda. Metode ini merupakan metode kuadrat terkecil, perbedaannya pada metode regresi *ridge* ini adalah nilai variabel prediktornya ditransformasikan dahulu melalui prosedur *centering* dan *rescaling* dengan menambah tetapan bias k yang relatif kecil pada diagonal utama matriks $\boldsymbol{X}^T\boldsymbol{X}$, sehingga persamaan model regresi menjadi:

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta}_* + \boldsymbol{\varepsilon}\\ \boldsymbol{y} &= (\boldsymbol{XT})(\boldsymbol{T}^{\boldsymbol{T}}\boldsymbol{\beta}_*) + \boldsymbol{\varepsilon}\\ \boldsymbol{y} &= \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},\end{aligned} \tag{1}$$

dengan $\boldsymbol{Z} = \boldsymbol{XT}$ dan $\boldsymbol{\alpha} = \boldsymbol{T}^T\boldsymbol{\beta}_*$
(Montgomery dan Peck, 1992)

Selanjutnya, parameter dalam regresi *ridge* $\boldsymbol{y} = \boldsymbol{Z}\widehat{\boldsymbol{\alpha}} + \boldsymbol{\varepsilon}$**,** diestimasi dengan metode *Ordinary Least Square* (OLS) yang meminimumkan jumlah kuadrat *error*. Estimator untuk regresi *ridge* adalah:

$$\widehat{\boldsymbol{\alpha}}_{RR} = (\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-\boldsymbol{1}}\boldsymbol{Z}^T\boldsymbol{y} \tag{2}$$

dengan $0 \leq k \leq \infty$. Estimator regresi *ridge* pada persamaan (2) dapat dituliskan sebagai berikut;

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{RR} &= (\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-\boldsymbol{1}}\boldsymbol{Z}^T\boldsymbol{y}\\ &= [\boldsymbol{I} - k(\boldsymbol{Z}^T\boldsymbol{Z} + k\boldsymbol{I})^{-\boldsymbol{1}}](\boldsymbol{Z}^T\boldsymbol{Z})^{-\boldsymbol{1}}\boldsymbol{Z}^T\boldsymbol{y}\end{aligned}$$

dimana telah diketahui bahwa $\boldsymbol{Z}^T\boldsymbol{Z} = \boldsymbol{\Lambda}$ dan $\widehat{\boldsymbol{\alpha}}_{LS} = \boldsymbol{\Lambda}^{-\boldsymbol{1}}\boldsymbol{Z}^T\boldsymbol{y}$, sehingga

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{RR} &= [\boldsymbol{I} - k(\boldsymbol{\Lambda} + k\boldsymbol{I})^{-\boldsymbol{1}}]\boldsymbol{\Lambda}^{-\boldsymbol{1}}\boldsymbol{Z}^T\boldsymbol{y}\\ &= [\boldsymbol{I} - k(\boldsymbol{\Lambda} + k\boldsymbol{I})^{-\boldsymbol{1}}]\,\widehat{\boldsymbol{\alpha}}_{LS}\end{aligned} \tag{3}$$

Estimasi regresi *ridge* dari $\boldsymbol{\beta}$ ialah : $\widehat{\boldsymbol{\beta}}_{RR} = \boldsymbol{T}\widehat{\boldsymbol{\alpha}}_{RR}$**,** di mana estimasi RR memiliki nilai MSE sebesar:

$$\begin{aligned}\text{MSE }(\widehat{\boldsymbol{\alpha}}_{RR}) &= Var(\widehat{\boldsymbol{\alpha}}_{RR}) + [Bias(\widehat{\boldsymbol{\alpha}}_{RR})]^2\\ &= \hat{\sigma}^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i + k_i)^2} + k^2\sum_{i=1}^{p}\frac{\hat{\alpha}_i^2}{(\lambda_i + k_i)^2}\\ &= \sum_{i=1}^{p}\frac{\hat{\sigma}^2\lambda_i + {k_i}^2\hat{\alpha}_i^2}{(\lambda_i + k_i)^2}\end{aligned} \tag{4}$$

Ridge Regression (RR) yang diperkenalkan oleh Hoerl dan Kennard pada tahun 1970 telah banyak digunakan oleh peneliti untuk menangani kasus multikolinieritas. Meskipun RR dapat mengatasi kasus multikolinieritas dengan nilai MSE yang lebih kecil daripada metode OLS. Namun estimator RR memiliki kelemahan, yakni ketika konstanta bias k telah terpilih maka belum dapat menghasilkan formula regresi *ridge* secara ekspilist dan tunggal. Salah satu cara penenganan multikolinieritas adalah dengan menggunakan regresi *ridge*. Namun, terdapat banyak pemilihan nilai parameter k dalam regresi ridge. Sampai saat ini terdapat sekitar 170 metode pemilihan nilai parameter k, sehingga menjadikan solusi regresi *ridge* menjadi tidak tunggal (Dorugade, 2016).

Dilakukan modifikasi penentuan konstanta *k* regresi *ridge* dengan ide awalnya ialah nilai koefisien korelasi antara *regressor* dapat membantu mendeteksi ketergantungan linier yang dekat atau multikolinieritas. Rodgers dan Nicewander (1988) menyampaikan cara untuk menginterpretasikan koefisien korelasi, dengan menyajikan 13 cara berbeda dari interpretasi untuk konsep koefisien korelasi sebagai ukuran paling dasar dari hubungan bivariat. Berdasarkan salah satu penafsiran ini, untuk data *bivariate* (X, Y) ketika distandarisasi dua variabel baku, standar deviasi nilainya menjadi 1 dan kemiringan garis regresi atau koefisien regresi sama dengan nilai koefisien korelasi antara Y dan X. Dalam hal ini konstan adalah 0 dan garis regresi dapat dinyatakan $\hat{Y} = rX$.

Nilai korelasi antara Y dan X dapat dinyatakan dalam bentuk matriks $\boldsymbol{X^T y}$. Maka dengan mengambil ide tersebut, Dorugade (2016) memodifikasi estimator dari RR dengan mengganti konstanta bias k dengan memanfaatkan nilai dari matriks $\boldsymbol{Z^T y}$. Sehingga, diperoleh estimator baru yang disebut dengan *Adjusted Ridge Regression* (AR) estimator dengan matriks diagonal $\boldsymbol{C}$ dan meminimumkan nilai bias dengan menggunakan fungsi akar. dengan matriks $\boldsymbol{Z^T y}$,

$$\boldsymbol{Z^T y} = \begin{bmatrix} Z_{11} & Z_{21} & \cdots & Z_{n1} \\ Z_{12} & Z_{22} & \cdots & Z_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{1p} & Z_{2p} & \cdots & Z_{np} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Z_{i1} y \\ \sum_{i=1}^{n} Z_{i2} y \\ \vdots \\ \sum_{i=1}^{n} Z_{ip} y \end{bmatrix}$$

dimana $r_{zy} = \frac{\sum_{i=1}^{n} Z_{ij} y}{\sqrt{\sum_{i=1}^{n} Z_{ij}{}^2}\sqrt{\sum_{i=1}^{n} y_i{}^2}}$, dengan nilai $\sqrt{\sum_{i=1}^{n} y_i{}^2} = 1$ (standarisasi)

$$\sum_{i=1}^{n} Z_{ij} y = r_{zy} \sqrt{\sum_{i=1}^{n} Z_{ij}{}^2}$$

Berdasarkan persamaan (3) maka estimator *AR* dapat dituliskan:

$$\hat{\boldsymbol{\alpha}}_{AR} = (\boldsymbol{\Lambda} + \boldsymbol{C})^{-\boldsymbol{1}} \boldsymbol{Z^T y}$$

dimana $C = diagonal\ [(|\boldsymbol{Z^T y}|)^{\frac{1}{2}}]$, atau

$$\hat{\boldsymbol{\alpha}}_{AR} = [\boldsymbol{I} - \boldsymbol{C}\boldsymbol{A}^{-\boldsymbol{1}}]\, \hat{\boldsymbol{\alpha}}_{\boldsymbol{LS}} \tag{5}$$

dimana $\boldsymbol{A} = \boldsymbol{\Lambda} + \boldsymbol{C}$

Estimator AR dari $\boldsymbol{\beta}$ ialah : $\hat{\boldsymbol{\beta}}_{AR} = \boldsymbol{T}\hat{\boldsymbol{\alpha}}_{AR}$, di mana nilai MSE dari AR adalah:

$$\begin{aligned} \text{MSE}\ (\hat{\boldsymbol{\alpha}}_{AR}) &= Var\ (\hat{\boldsymbol{\alpha}}_{\boldsymbol{AR}}) + [Bias(\hat{\boldsymbol{\alpha}}_{\boldsymbol{AR}})]^{\boldsymbol{2}} \\ &= \sum_{i=1}^{p} \frac{\sigma^2 \lambda_i}{(\lambda_i + c_i)^2} + \sum_{i=1}^{p} \frac{(c_i \alpha_i)^2}{(\lambda_i + c_i)^2} \\ &= \sum_{i=1}^{p} \frac{\sigma^2\, \lambda_i + (\alpha_i c_i)^2}{(\lambda_i + c_i)^{\,2}} \end{aligned} \tag{6}$$

dimana c_i merupakan elemen ke- i dari diagonal C, dan $i = 1,2, .., p$. Adapun $\hat{\alpha}_i$ merupakan elemen estimator dari $\hat{\alpha}_{LS}$ dengan $i = 1,2, \ldots p$ dan $\hat{\sigma}^2$adalah MSE dari estimator $\hat{\sigma}^2 = \frac{Y^T Y - \hat{\alpha}_{LS}{}^T Z^T Y}{n-p-1}$

3. METODE PENELITIAN

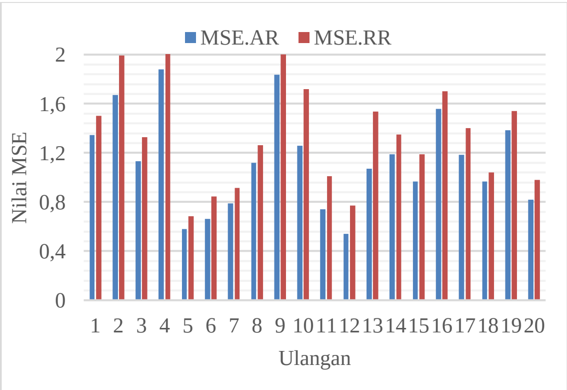
Analisis *Ridge Regression* (RR) dan *Adjusted Ridge Regression* (AR) dilakukan dengan menggunakan simulasi data. Data simulasi yang dibangkitkan merupakan data normal multivariat yang diulang sebanyak 20 kali dengan beberapa ukuran sampel dan ukuran nilai korelasi. Ukuran sampel digunakan ialah 20, 50, dan 100 data. Adapun ukuran nilai korelasi yang digunakan ialah 0.6, 0.8, dan 0.9. Kombinasi dari data simulasi tersebut menunjukkan kondisi data dengan masalah multikolinearitas rendah, sedang dan tinggi. Pembangkitan data untuk setiap ukuran sampel dilakukan dengan program *RStudio* menggunakan perintah *mvtnorm* dengan fungsi *rmvnorm* untuk membangkitkan data variabel predictor dan fungsi *rnorm* untuk membangkitkan variabel respon.

Metode analisis yang dilakukan dalam analisis perbandingan metode AR dan RR dalam menangani kasus multikolinearitas mengikuti langkah-langkah sebagai berikut:

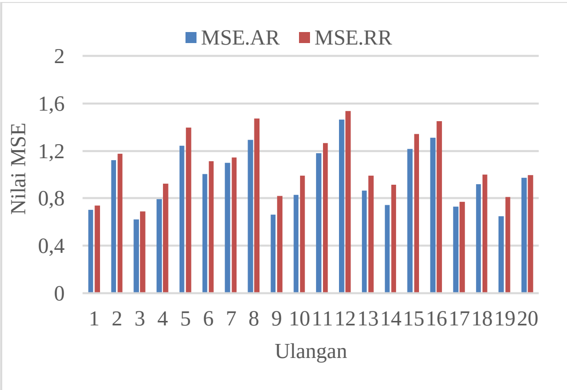
1. Membangkitkan data dengan kombinasi ukuran sampel dan korelasi yang telah ditentukan sebelumnya.
2. Perhitungan nilai MSE untuk setiap metode, yaitu menggunakan metode *Ridge Regression* (RR) dan *Adjusted Ridge Regression* (AR).
3. Eksplorasi nilai MSE pada masing-masing metode dan kombinasi data.
4. Analisis perbandingan nilai MSE antara metode AR dan RR berdasarkan ukuran sampel dan besarnya masalah multikolinearitas.
5. Penarikan kesimpulan berdasarkan hasil analisis data simulasi

4. HASIL DAN PEMBAHASAN

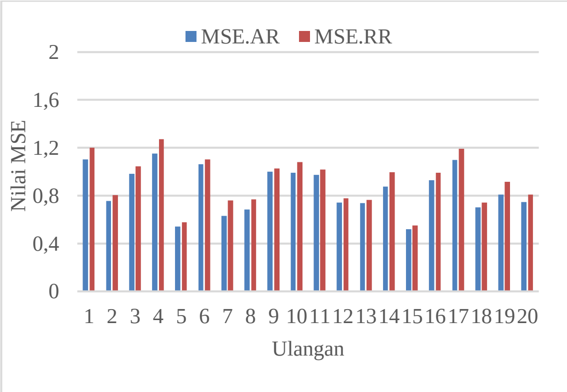
Berikut merupakan hasil perhitungan nilai MSE dari masing-masing metode yakni *Ridge Regression* (RR) dan *Adjusted Ridge Regression* (AR) pada ukuran sampel $n = 20$ dengan masing-masing ukuran korelasi yaitu 0.6, 0.8, dan 0.9. Hasil perhitungan tersebut disajikan dalam bentuk grafik sebagai berikut:



Gambar 1a. Grafik MSE AR dan MSE RR $(n = 20\,;\ \rho = 0.6)$

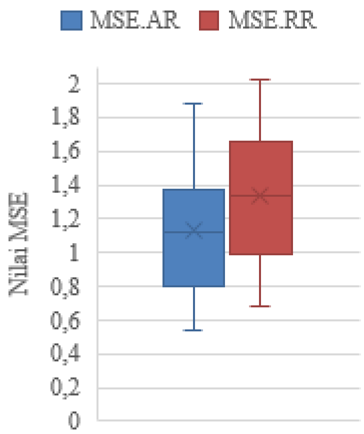


Gambar 1b. Grafik MSE AR dan MSE RR $(n = 20\,;\ \rho = 0.8)$

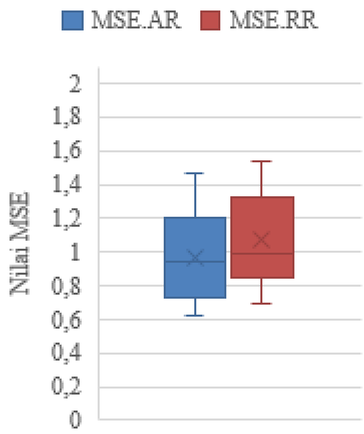


Gambar 1c. Grafik MSE AR dan MSE RR ($n = 20\ ;\ \rho = 0.9$)

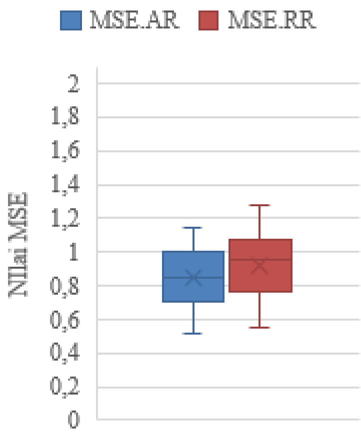
Berdasarkan keseluruhan Gambar 1 di atas, dapat dilihat bahwa nilai MSE AR lebih kecil dibandingkan nilai MSE RR. Hal ini ditunjukkan dengan diagram batang pada MSE AR yang berukuran lebih pendek dibandingkan dengan MSE RR di setiap ulangan. Selain itu, dapat dailihat pula bahwa nilai MSE semakin menurun seiring meningkatnya ukuran korelasi baik pada AR mapun RR. Lebih jauh, sebaran nilai MSE dari AR maupun AR dapat dilihat pada *boxplot* sebagai berikut:



Gambar 2a. *Boxplot* MSE AR dan MSE RR ($n = 20\ ;\ \rho = 0.6$)

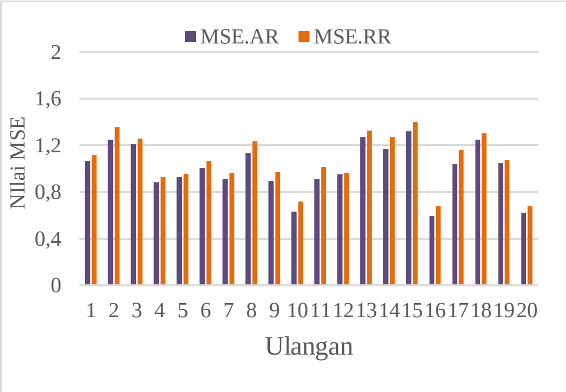


Gambar 2b. *Boxplot* MSE AR dan MSE RR ($n = 20\ ;\ \rho = 0.8$)

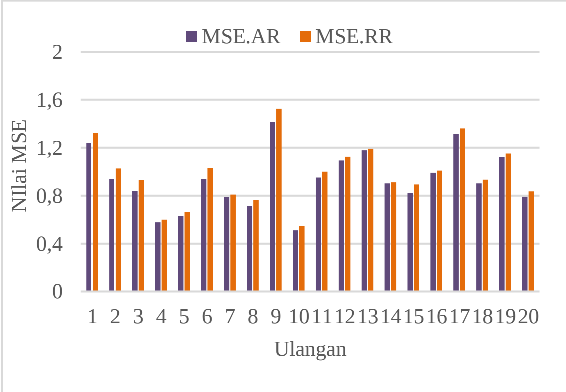


Gambar 2c. *Boxplot* MSE AR dan MSE RR ($n = 20\ ;\ \rho = 0.9$)

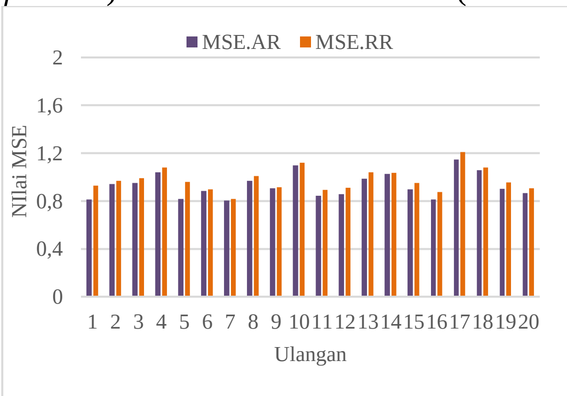
Berdasarkan Gambar 2 di atas, dapat dilihat bahwa *boxplot* MSE dari AR maupun RR cenderung mengecil ketika ukuran korelasi meningkat pada ukuran sampel yang kecil. Hal ini menunjukkan bahwa keragaman MSE semakin mengecil dan cenderung lebih stabil seiring dengan meningkatnya nilai korelasi. Selain itu, dapat dilihat pula bahwa nilai MSE AR memiliki *range*/rentang serta median yang lebih kecil dibandingkan nilai MSE RR pada setiap ukuran korelasi yaitu $\rho = 0.6$, $\rho = 0.8$, dan $\rho = 0.9$ untuk ukuran sampel $n = 20$. Hal ini menunjukkan bahwa metode AR lebih baik dalam menangani masalah multikolinearitas dibandingkan metode RR pada ukuran sampel kecil. Untuk melihat lebih jauh pengaruh ukuran korelasi terhadap nilai MSE dari kedua metode, maka dilakukan analisis menggunakan ukuran sampel yang lebih besar yaitu $n = 50$ sebagai berikut:



Gambar 3a. Grafik MSE AR dan MSE RR ($n = 50\ ;\ \rho = 0.6$)

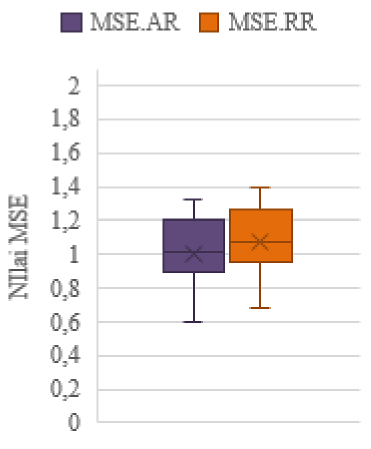


Gambar 3b. Grafik MSE AR dan MSE RR ($n = 50\ ;\ \rho = 0.8$)

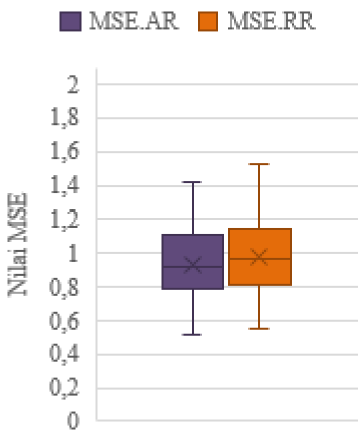


Gambar 3c. Grafik MSE AR dan MSE RR ($n = 50\ ;\ \rho = 0.9$)

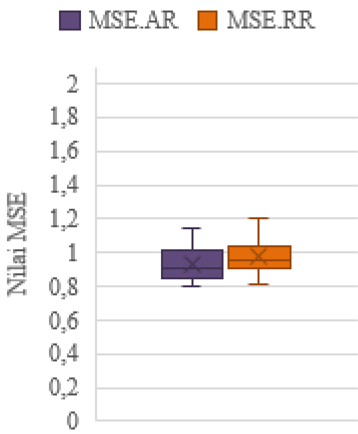
Berdasarkan Gambar 3 di atas, dapat dilihat bahwa pada ukuran sampel $n = 50$, nilai MSE cenderung mengalami penuruan ketika ukuran korelasi semakin meningkat baik pada metode AR maupun RR. Penurunan nilai MSE tersebut cukup signifikan dan konstan di setiap ulangan pada ukuran korelasi tertinggi yaitu $\rho = 0.9$. Sedangkan pada ukuran korelasi $\rho = 0.8$, penurunan nilai MSE tidak begitu signifikan terlihat terhadap ukuran korelasi $\rho = 0.6$. Berdasarkan keseluruhan grafik pada Gambar 3 di atas, dapat dilihat bahwa nilai MSE AR lebih kecil dibandingkan nilai MSE RR di setiap ulangan pada setiap ukuran korelasi.



Gambar 4a. *Boxplot* MSE AR dan MSE RR ($n = 50\ ;\ \rho = 0.6$)

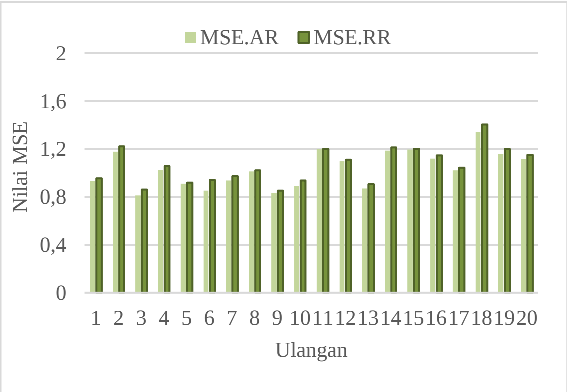


Gambar 4b. *Boxplot* MSE AR dan MSE RR ($n = 50\ ;\ \rho = 0.8$)

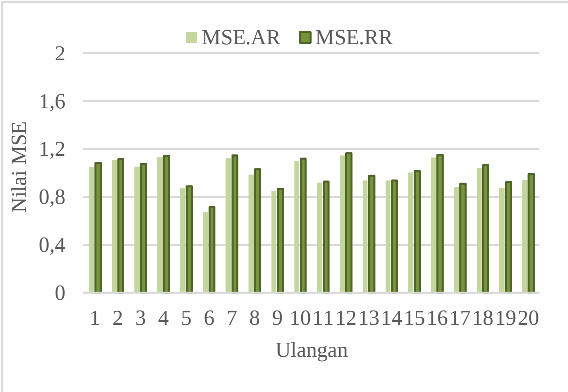


Gambar 4c. *Boxplot* MSE AR dan MSE RR ($n = 50\ ;\ \rho = 0.9$)

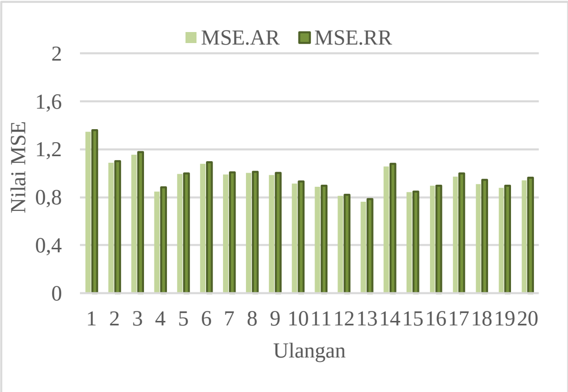
Berdasarkan Gambar 4 di atas, dapat dilihat bahwa pada ukuran sampel $n = 50$, *boxplot* MSE dari AR maupun RR cenderung mengecil seiring meningkatnya ukuran korelasi. *Boxplot* terkecil diperoleh pada ukuran korelasi $\rho = 0.9$ yang ditunjukkan pada Gambar 4c dengan median yang lebih mendekati nol dibandingkan ukuran korelasi lainnya. Hal ini menunjukkan bahwa nilai MSE pada ukuran korelasi tertinggi cenderung stabil dan konstan mendekati nol dengan keragaman yang kecil di setiap ulangan. Adapun median dari nilai MSE pada ukuran korelasi $\rho = 0.8$ juga terlihat lebih kecil dibandingkan ukuran korelasi $\rho = 0.6$ meskipun *range boxplot* nilai MSE pada ukuran korelasi $\rho = 0.8$ terlihat lebih besar dibandingkan pada ukuran korelasi yang lebih kecil yaitu $\rho = 0.6$. Selain itu, dapat dilihat pula bahwa median nilai MSE dari metose AR konstan lebih kecil dibandingkan nilai MSE dari metode RR pada setiap ukuran korelasi.



Gambar 5a. Grafik MSE AR dan MSE RR ($n = 100$; $\rho = 0.6$)

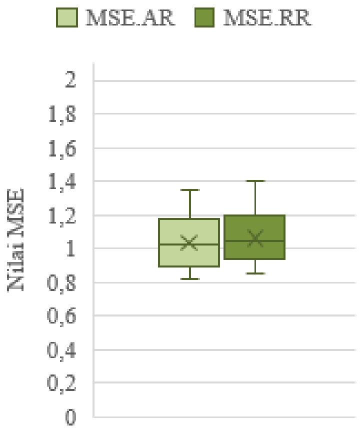


Gambar 5b. Grafik MSE AR dan MSE RR ($n = 100$; $\rho = 0.8$)

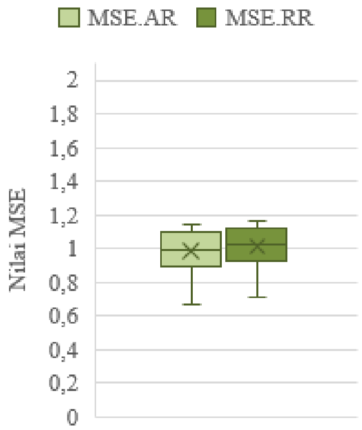


Gambar 5c. Grafik MSE AR dan MSE RR ($n = 100$; $\rho = 0.9$)

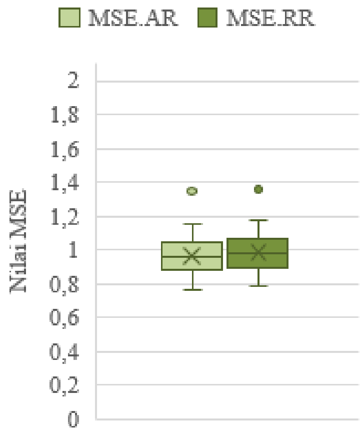
Berdasarkan Gambar 5 di atas, dapat dilihat bahwa pada ukuran sampel $n = 100$, nilai MSE cenderung mengalami penurunan ketika ukuran korelasi semakin meningkat baik pada metode AR maupun RR. Nilai MSE pada ukuran korelasi $\rho = 0.6$ terlihat tidak jauh berbeda dengan nilai MSE pada ukuran korelasi $\rho = 0.8$. Namun demikian, nilai MSE tersebut menurun cukup signifikan pada ukuran korelasi yang lebih besar yaitu $\rho = 0.9$. Selain itu, dapat dilihat pula bahwa nilai MSE metode AR lebih kecil dibandingkan metode RR pada setiap ukuran korelasi maupun setiap ulangan.



Gambar 6a. *Boxplot* MSE AR dan MSE RR ($n = 100\ ;\ \rho = 0.6$)

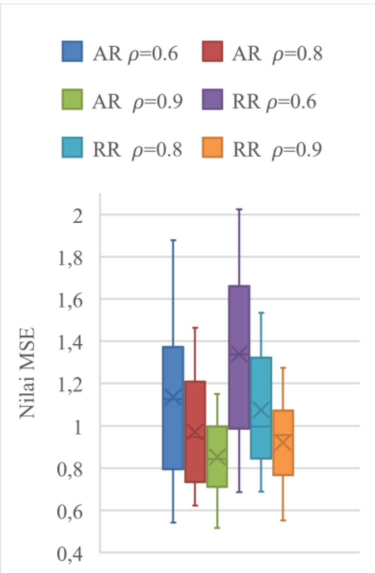


Gambar 6b. *Boxplot* MSE AR dan MSE RR ($n = 100\ ;\ \rho = 0.8$)

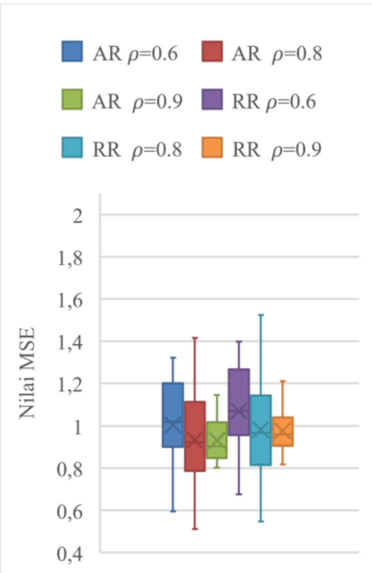


Gambar 6c. *Boxplot* MSE AR dan MSE RR ($n = 100\ ;\ \rho = 0.9$)

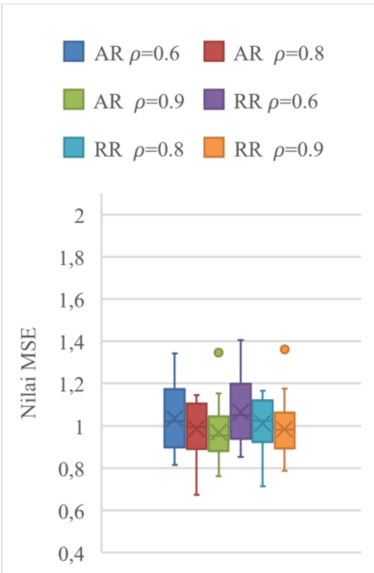
Berdasarkan Gambar 6 di atas, dapat dilihat bahwa pada ukuran sampel $n = 100$, *boxplot* nilai MSE dari metode AR maupun RR cenderung mengecil seiring meningkatnya ukuran korelasi. Selain itu, dapat pula dilihat bahwa nilai median cenderung mengalami penurunan ketika ukuran korelasi semakin menngkat. *Range* dari *boxplot* juga semakin kecil seiring bertambahnya ukuran korelasi meskipun terlihat adanya pencilan pada ukuran korelasi tertinggi yaitu $\rho = 0.9$. Hal ini menunjukkan bahwa pada ukuran sampel cukup besar, yaitu $n = 100$, metode AR maupun metode RR akan semakin baik digunakan dalam mengatasi masalah multikolinearitas ketika ukuran korelasi antar variabel prediktor mendekati satu. Lebih jauh, dapat dilihat pula bahwa nilai MSE dari AR selalu lebih kecil dibandingkan nilai MSE RR di setiap ukuran korelasi. Hal ini mengindikasikan bahwa metode AR lebih baik dibandingkan metode RR dalam mengatasi masalah multikolinearitas meskipun perbandingan nilai MSE yang dihasilkan tidak begitu jauh pada ukuran sampel ini. Sebagai bahan analisis yang lebih dalam, berikut disajikan *boxplot* perbandingan nilai MSE antara AR dan RR pada ukuran korelasi berbeda dengan ukuran sampel tetap.



Gambar 7a. *Boxplot* MSE AR dan RR pada $n = 20$



Gambar 7b. *Boxplot* MSE AR dan RR pada $n = 50$

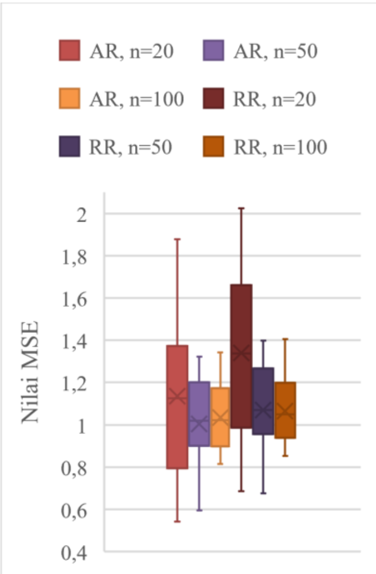


Gambar 7c. *Boxplot* MSE AR dan RR pada $n = 100$

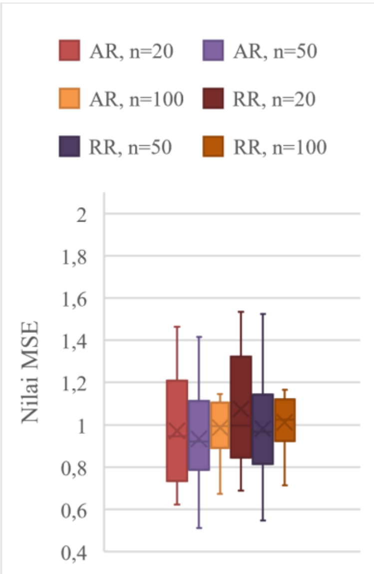
Berdasarkan Gambar 7 di atas, dapat dilihat bahwa ukuran sampel cukup memberikan pengaruh dalam penurunan nilai MSE pada kedua metode. Hal ini ditunjukkan dengan semakin kecilnya ukuran *boxplot* seiring dengan bertambahnya sampel yang digunakan. Selain itu, dapat dilihat pula bahwa *range boxplot* semakin mengecil seiring bertambahnya ukuran sampel. Hal ini menunjukkan pula bahwa nilai MSE cenderung semakin stabil pada ukuran sampel lebih besar dengan ragam nilai MSE yang relatif rendah.

Perbandingan kedua metode yaitu AR dan RR pada ukuran sampel yang lebih kecil terlihat cukup signifikan. Hal ini terlihat pada Gambar 7a di mana metode AR menghasilkan nilai MSE yang lebih kecil dibadingkan metode RR untuk setiap ukuran korelasi. Hal ini juga terlihat pada ukuran sampel yang lebih besar di mana nilai MSE metode AR lebih kecil dibandingkan RR meskipun dengan selisih yang relatif kecil. Hal ini ditunjukkan pula dengan posisi median *boxplot* yang tidak jauh berbeda pada ukuran sampel besar yaitu $n = 50$ dan $n = 100$ namun berbeda signifikan pada ukuran sampel kecil yaitu $n = 20$.

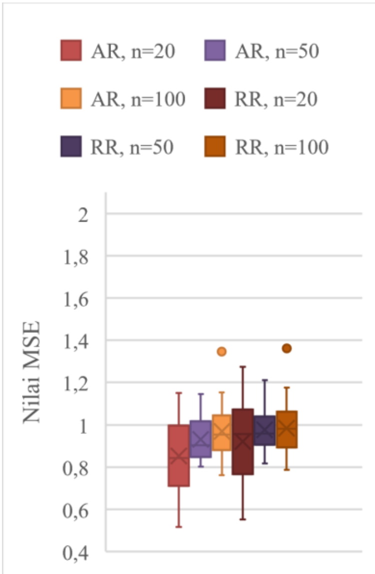
Sebagai bahan perbandingan lain, disajikan *boxplot* nilai MSE dari masing-masing metode untuk ukuran sampel berbeda dengan ukuran korelasi tetap sebagai berikut:



Gambar 8a. *Boxplot* MSE AR dan RR pada $\rho = 0{,}6$



Gambar 8b. *Boxplot* MSE AR dan RR pada $\rho = 0{,}8$



Gambar 8c. *Boxplot* MSE AR dan RR pada $\rho = 0{,}9$

Berdasarkan Gambar 8 di atas, dapat dilihat bahwa ukuran korelasi cukup memberikan pengaruh dalam *range* dan median *boxplot* nilai MSE pada kedua metode. Hal ini ditunjukkan dengan semakin kecilnya *range boxplot* serta semakin rendahnya nilai median seiring dengan meningkatnya ukuran korelasi. Pada ukuran korelasi yang relatif kecil yaitu $\rho = 0.6$, perbedaan ukuran *boxplot* cukup terlihat antara ukuran sampel kecil $n = 20$ dan sampel besar $n = 100$. Sedangkan pada ukuran korelasi yang lebih besar yaitu $\rho = 0.8$, *boxplot* menunjukkan ukuran yang cenderung mengecil seiring meningkatnya ukuran sampel. Begitu pula pada ukuran korelasi terbesar yaitu $\rho = 0.9$, *boxplot* terlihat mengecil dengan bertambahnya ukuran sampel. Namun demikian, pada ukuran korelasi tersebut, perbedaan nilai MSE yang dihasilkan tidak begitu signifikan antara ukuran sampel $n = 50$ dan $n = 100$ yang ditunjukkan dengan ukuran *boxplot* yang hampir sama dan posisi median yang hampir sejajar. Perbandingan kedua metode yaitu AR dan RR pada setiap ukuran korelasi yaitu $\rho = 0.6$, $\rho = 0{,}8$, dan $\rho = 0{,}9$ cukup signifikan terutama pada ukuran sampel kecil. Hal ini menunjukkan bahwa metode AR cenderung lebih baik digunakan dalam mengatasi masalah multikolinieritas yang terjadi antar variabel prediktor.

5. KESIMPULAN

Berdasarkan kajian simulasi yang telah diuraikan di atas, dapat ditarik kesimpulan bahwa metode AR dan metode RR menghasilkan nilai MSE yang lebih kecil ketika ukuran sampel yang digunakan semakin besar. Nilai MSE yang dihasilkan oleh metode AR cenderung lebih kecil dibandingan dengan metode RR yang terlihat dari setiap ulangan data yang digunakan. Hal ini menunjukkan bahwa metode AR relatif lebih efektif dibandingkan dengan metode RR. Hal ini senada dengan simulasi yang dilakukan oleh Dorugade (2016), bahwa kecenderungan AR lebih efektif untuk menangani kasus multikolinearitas dibandingkan metode RR.

DAFTAR PUSTAKA

Daoud, J.I., 2017. Multicollinearity and Regression Analysis. J. Phys. Conf. Ser. 949, 012009. https://doi.org/10.1088/1742-6596/949/1/012009

Dorugade, A.V., 2016. *Adjusted ridge estimator and comparison with Kibria's method in linear regression*. J. Assoc. Arab Univ. Basic Appl. Sci. 21, 96–102. https://doi.org/10.1016/j.jaubas.2015.04.002

Hildawati, H., Rusgiyono, A., Sudarno, S., 2016. Pemodelan Upah Minimum Kabupaten/Kota di Jawa Tengah Berdasarkan Faktor-Faktor yang Mempengaruhinya Menggunakan Regresi Ridge. *J. Gaussian 5*, 123–132. https://doi.org/10.14710/j.gauss.5.1.123-132

Hoerl, A.E., Kennard, R.W., 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* 12, 55–67. https://doi.org/10.1080/00401706.1970.10488634

Iqbal, M.A., 2021. *Application of Regression Techniques with their Advantages and Disadvantages* 4, 11–17.

Iriawan, N., Astuti, S.P., 2006. *Mengolah data statistik dengan mudah menggunakan minitab 14*. Yogyak. Andi.

Montgomery, D.C., Peck, E.A., 1992. *Introduction to Linear Regression Analysis*, Wiley Series in Probability and Statistics - Applied Probability and Statistics Section. Wiley.

Rodgers, J.L., Nicewander, W.A., 1988. *Thirteen Ways to Look at the Correlation Coefficient.* Am. Stat. 42, 59–66. https://doi.org/10.2307/2685263

Sulistianingsih, E., Suparti, S., Ispriyanti, D., 2023. Pemodelan Indeks Pembangunan Manusia di Jawa Tengah Menggunakan Metode Regresi Ridge dan Regresi Stepwise. *J. Gaussian 11*, 468–477. https://doi.org/10.14710/j.gauss.11.3.468-477

Tazliqoh, A.Z., Rahmawati, R., Safitri, D., 2015. Perbandingan Regresi Komponen Utama dengan Regresi Ridge pada Analisis Faktor-Faktor Pendapatan Asli Daerah (PAD) Provinsi Jawa Tengah. *J. Gaussian 4*, 1–10. https://doi.org/10.14710/j.gauss.4.1.1-10